

DOCUMENT RESUME

ED 268 010

SE 046 538

AUTHOR Fisher, Kathleen M.
TITLE Systematic Representation of Molecular Biology Knowledge.
PUB DATE Oct 85
NOTE 6p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (59th, San Francisco, CA, March 28-April 1, 1986). For related papers presented in the same set, see SE 046 537-540.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Artificial Intelligence; *Biology; *Cognitive Psychology; *College Science; *Concept Formation; Higher Education; *Knowledge Level; Science Education; *Science Instruction
IDENTIFIERS Science Education Research

ABSTRACT

A small set of relationships has been identified which appears to be sufficient for describing all molecular and cellular reactions and structures discussed in an introductory biology course. A precise definition has been developed for each relationship. These 20 relationships are of four types: (1) analytical; (2) spatial; (3) temporal; and (4) process. Three different computer-based semantic network representations of molecular biology knowledge were developed. Two list-structure nets were useful in identifying problems in net design. The most successful of the three prototypes involved spatial representations (concept maps) for each concept, with the maps being linked together to form a network. Various styles of knowledge representation were tested with students enrolled in introductory biology. The students were most enthusiastic about exercises in which partially completed concept maps were given to them and they had to fill in the missing pieces (concepts or relationships). The systematic analysis of relationships between concepts reveals relationships not explicitly perceived before, even by experts. It is believed that these representational techniques will be useful in facilitating student learning. (Author/JN)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED268010

II. Systematic Representation of Molecular Biology Knowledge

Kathleen M. Fisher

Department of Education &
Division of Biological Sciences
University of California — Davis
Davis, CA 95616

October 1985

Abstract

The first paper in this set outlines the rationale for developing systematic means of representation of biological knowledge using techniques from artificial intelligence and cognitive science. This paper describes the analysis of one subdomain, molecular biology, in the context of an introductory biology course.

A small set of relationships has been identified which appears to be sufficient for describing all molecular and cellular reactions and structures discussed in an introductory biology course. A precise definition has been developed for each relationship. These twenty relationships are of four types: analytical (e.g., a "set" and its "members"), spatial (e.g., "contains" and "contained within"), temporal (e.g., "precedes" and "follows"), and process (e.g., "input," "agent").

Three different computer-based semantic network representations of molecular biology knowledge were developed. Two list-structure nets were useful in identifying problems in net design. The most successful of the three prototypes involved spatial representations (concepts maps) for each concept, with the maps being linked together to form a network (details are described in the third paper of the set).

Various styles of knowledge representation were tested with students enrolled in introductory biology. The students were most enthusiastic about exercises in which partially completed concept maps were given to them and they had to fill in the missing pieces (concepts or relationships).

The systematic analysis of relationships between concepts reveals relationships not explicitly perceived before, even by experts. We believe that these representational techniques will be useful in facilitating student learning.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

* Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Kathleen M. Fisher

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

SE046538

II. Systematic Representation of Molecular Biology Knowledge

Kathleen M. Fisher

Department of Education &
Division of Biological Sciences
University of California — Davis
Davis, CA 95616

October 1985

1. Purpose of Study

The purpose of this work is to assess the validity of the theory presented in the first paper. That is, can biological knowledge be practically represented in systematic ways using techniques drawn from artificial intelligence and cognitive science? And if so, what are the advantages and pitfalls of such representations? More specifically, the aims are: To identify a small set of relationships for describing one subdomain of biology, molecular biology, and to define these relationships precisely, to develop methods for utilizing those relationships in knowledge representation that seem attractive and useful (clear, concise, informative) to experts and novices alike, and to discover optimum ways of incorporating these representational methods into the instructional process.

2. Theoretical Basis of Study

The underlying rationale has been largely outlined in the first paper. Several additional considerations are summarized here. First, the choice of representation can be a crucial factor in determining success or failure in learning or problem solving. It seems highly unlikely that the forms of representation of biological knowledge that have evolved historically are the optimum forms for teaching biology or for solving problems in biology.

Second, the principal focus is on promoting *meaningful* rather than rote learning by biology students. Since understanding of a concept implies knowing how that concept is related to many other concepts, the aim is to make relationships between concepts explicit and to give students lots of practice in using them.

Third, the aim is not to create a unit of instruction, since students will be introduced to concepts in the usual ways (lectures, textbooks, films, laboratory). This project aims to create a microworld in which students can "play," thereby enriching their understanding of biology. More specifically, the envisioned computer-based semantic network aims to approximate the organization of biological knowledge in the mind of an expert, and at the same time encourage students to climb in and explore the knowledge.

Fourth, the two goals, (a) creating an approximation of the organization of biological knowledge in the mind of an expert and (b) creating a knowledge base which is informative and useful for undergraduate biology students, are often incompatible. When conflicts arise, the second consideration takes precedence.

Fifth, exploration of the network might not seem very attractive or motivating as an end in itself. The intention, therefore, is to create a number of games as overlays on the computer-based semantic network to provide an incentive for students to play in the knowledge base.

Finally, it is clear that each subdomain of biology involves different kinds of relationships. Interactions between molecules have some characteristics in common with interactions between species, but there are different aspects as well. We chose molecular biology as the simplest subdomain in which to begin the task of analysis. Its advantages include the following. It is relatively free of "historical baggage" and relatively precise, concise, and concrete. Also, there is a high degree of consensus among biologists concerning the nature of concepts (meanings of terms).

3. Procedures and Results of Study

The SEMNET group analyzed the knowledge structure of molecular biology and explored alternative forms of representation. The group eventually settled on four classes of relationships, three of which are fairly common: (1) *analytical* (such as a "set" and its "members", seven relationships), (2) *spatial* (such as "contains" and "contained in", 2 relationships), and (3) *temporal* (e.g., "precedes" and "follows", 3 relationships). The fourth set of relationships, (4) *process* relationships, are more novel. They involve identifying the role played by each component in a process (e.g., "input," "output," "informant," "agent") or describing the state of the system (8 relationships).

Two prototype semantic network structures were developed in computers, one on an Apple II-E and another on a Sun computer. Both prototypes involved list structures as means for both entering the data and for displaying the data on the screen. These prototypes allowed us to make two important observations. First, the list structures were very unsatisfying (difficult, cumbersome). Biologists who were creating data banks found they could not create a satisfactory network, because of the difficulties in thinking in lists. Secondly, we concluded that introducing synonyms for each relationship, while attractive in theory, added unnecessary confusion and ambiguity. We decided to aim for the most parsimonious list of precisely defined relationships we could generate.

Experimentation with a Macintosh computer led us to discover the virtues of concept mapping. In our hands this technique involves placing a concept to be described in the center of the page and connecting closely related concepts to it by labeled lines (relationships) as illustrated in Figure 1. With these spatial strategies, we could organize information on a screen or page in such a way that it was relatively easy to read and comprehend. We developed a set of rules for standardizing the concept mapping format. This standardization made it much easier to extract information from any given map.

A graduate student, M. J. Smith, spent a quarter working with a group of undergraduate students enrolled in an introductory biology course. Her aim was to determine what kinds of representations the students found most useful. One exercise, looking at pairs of components involved in a given process (such as mRNA and tRNA in translation) and identifying the role played by each, was extremely difficult for students. Another exercise given early in the quarter involved students in

generating their own concept maps. They did this with considerable difficulty working alone, and when they worked in groups, the quality of the maps degenerated (a case where compromise did more harm than good). Finally, a series of exercises was developed using partially completed concept maps; students were to fill in the missing pieces of information. Students were very enthusiastic about these exercises. A number of students told us that they made xeroxed copies of the maps and shared them with friends; others reported using mapping techniques in other courses.

4. Conclusions

The systematic representation of biological knowledge described here seems reasonably useful. Many questions are yet to be answered. For example, should bidirectional relationships be described with two different names, depending on direction, for completeness, (such as "contains" and "contained in") or should a single name be used for parsimony (e.g., "contain")? Are other subdomains of biology such as ecology (Garb, Fisher and Faletti below) or organismal biology or animal behavior amenable to the same kind of analysis? Does the concept mapping representation interfere with characteristics that seem desirable in a larger computer-based network?

5. Implications for Teaching

This work has strong implications for teaching introductory biology. One impetus for this work came from the observation that many students fail to see connections between pockets of knowledge. For example, many don't understand that meiosis and independent assortment of unlinked genetic traits are intimately linked processes. The hope is that this alternative, non-linear representation of knowledge may promote greater recognition of these interconnections by undergraduate students. Each of us who has worked with the concept maps and networks, including various experts in biology, have perceived new links we had not explicitly recognized before. In several cases, we also discovered that we had subtle but different ways of thinking about concepts (as where one expert thought of a photosystem as a collection of processes and another as a collection of structures in which processes occur). We believe that the representational techniques have often clarified our thinking, and we expect them to do the same for students. This hypothesis has not yet been formally tested.

At the same time, we are cognizant of the difficulties of influencing teaching practices. Two ambitious analyses of biological knowledge, one based on Bloom's taxonomy of thinking skills (Klopfer, 1971) and the other on a systems analysis of the field (Miller, 1978) have, to our knowledge, been little used by teachers. Our work is quite theoretical and its perceived utility by biology teachers remains an open question.

6. References

- Klopfer, L. E. (1971). Evaluation of learning in science. In Bloom, B S., Hastings, J.T., and Madaus, G.F. Eds., *Handbook on Formative and Summative Evaluation of Student Learning*. San Francisco: McGraw-Hill.

Miller, J. G. (1978). *Living Systems*. San Francisco: McGraw-Hill.

BEST COPY AVAILABLE